CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2022-0044

Vol. **4**, No. 4, pp. 776-796
December 2023

The Double Regularization Method for Capacity Constrained Optimal Transport

Tianhao Wu[1,2], Qihao Cheng[1], Zihao Wang[3], Chaorui Zhang[2], Bo Bai[2], Zhongyi Huang[1] and Hao Wu[1,*]

[1] *Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China.*
[2] *Theory Lab, Center Research Institute, 2012 Labs, Huawei Technologies Co. Ltd., Hong Kong SAR, P.R. China.*
[3] *Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong SAR, P.R. China.*

Received 25 November 2022; Accepted 5 June 2023

Abstract. Capacity constrained optimal transport is a variant of optimal transport, which adds extra constraints on the set of feasible couplings in the original optimal transport problem to limit the mass transported between each pair of source and sink. Based on this setting, constrained optimal transport has numerous applications, e.g., finance, network flow. However, due to the large number of constraints in this problem, existing algorithms are both time-consuming and space-consuming. In this paper, inspired by entropic regularization for the classical optimal transport problem, we introduce a novel regularization term for capacity constrained optimal transport. The regularized problem naturally satisfies the capacity constraints and consequently makes it possible to analyze the duality. Unlike the matrix-vector multiplication in the alternate iteration scheme for solving classical optimal transport, in our algorithm, each alternate iteration step is to solve several single-variable equations. Fortunately, we find that each of these equations corresponds to a single-variable monotonic function, and we convert solving these equations into finding the unique zero point of each single-variable monotonic function with Newton's method. Theoretical analysis further provides a convergence guarantee to our algorithm. Extensive numerical experiments demonstrate that our proposed method has a significant advantage in terms of accuracy, efficiency, and memory consumption compared with existing methods.

AMS subject classifications: 49M25, 65K10

Key words: Capacity constraint, optimal transport, regularization, Sinkhorn algorithm.

*Corresponding author. *Email addresses:* wuth20@mails.tsinghua.edu.cn (T. Wu), cqh22@mails.tsinghua.edu.cn (Q. Cheng), zwanggc@cse.ust.hk (Z. Wang), chaorui.zhang@gmail.com (C. Zhang), ee.bobbai@gmail.com (B. Bai), zhongyih@tsinghua.edu.cn (Z. Huang), hwu@tsinghua.edu.cn (H. Wu).

1 Introduction

In this work, we present a novel time-efficient and space-saving method, to solve the discrete capacity constrained optimal transport (COT) problem [6], which is a discretization of continuous COT [29] and defined as

$$\begin{aligned} &\min_{\gamma\in\mathbb{R}_+^{N\times M}} \langle \boldsymbol{C},\boldsymbol{\gamma}\rangle, \\ &\text{s.t.}\quad \boldsymbol{\theta}\le\boldsymbol{\gamma}\le\boldsymbol{\eta},\quad \boldsymbol{\gamma}\mathbf{1}=\boldsymbol{u},\quad \boldsymbol{\gamma}^T\mathbf{1}=\boldsymbol{v}. \end{aligned} \tag{1.1}$$

Here, $\boldsymbol{\eta}=[\eta_{ij}]\in\mathbb{R}_+^{N\times M}$ and $\boldsymbol{\theta}=[\theta_{ij}]\in\mathbb{R}_+^{N\times M}$ are the given capacity upper and lower bound matrices, respectively. $\boldsymbol{u}$ and $\boldsymbol{v}$ are marginal distributions satisfying $\|\boldsymbol{u}\|_1=\|\boldsymbol{v}\|_1=1$. $\boldsymbol{\gamma}=[\gamma_{ij}]_{N\times M}$ is the transport plan, whose element γ_{ij} denotes the mass transported from position x_i to y_j, and is upper and lower bounded by η_{ij} and θ_{ij} in the COT problem, i.e., $\theta_{ij}\le\gamma_{ij}\le\eta_{ij}$. $\boldsymbol{A}\le\boldsymbol{B}$ denotes that every element a_{ij} in matrix $\boldsymbol{A}$ is not larger than corresponding b_{ij} in matrix $\boldsymbol{B}$. Note that when $\theta_{ij}=0$ and $\eta_{ij}=1$ for any $i=1,\dots,N$ and $j=1,\dots,M$, the discretized COT problem (1.1) is exactly the classical optimal transport problem, in which $\boldsymbol{C}=[c_{ij}]\in\mathbb{R}_+^{N\times M}$ denotes the cost matrix.†

Considering the capacity upper and lower bounds in the capacity constrained optimal transport problem, there are a large number of applications in various fields. For network flow [19], COT can be formulated as a minimum cost maximum flow problem [4,18,47], in which each edge has a capacity. For asset pricing and hedging in finance [1,14,44], the duality of martingale constrained optimal transport is closely associated with the fundamental theorem of asset pricing.

Optimal transport theory [22,27,28] has been successfully applied in different fields [8,9,17,21,30,38,40,41,50,52]. Consequently, there are numerous algorithms for solving classical optimal transport problem proposed in different perspectives, such as linear programming methods [51], primal-dual algorithms [23], solving Monge-Ampère equation [5,13,26,39], proximal block coordinate descent methods [24], reduction and approximation techiniques for high-dimensional distributions [31,35,36,53], Sinkhorn algorithm and its variants [2,3,12,32–34,42], etc. However, the large number of constraints added in the COT problem brings out new challenges for computation, and the methods mentioned above can hardly be applied directly to the COT problem. A straightforward solution is to formulate the COT problem as a min-cost-max-flow problem on a complete bipartite graph and then solve it with the network flow algorithms, which incurs cubic time complexity according to [48,49]. In addition, as pointed out in [15,34], solutions obtained by network flow algorithms are indifferentiable. There are also several algorithms [6,11,54] designed specifically for solving COT problems, among which iterative Bregman projection (IBP) proposed in [6] is the most common method. The main idea of IBP is to project the solution onto part of the constraint set alternately based on

†In this paper, our discussion is general for any N and M. For the sake of simplicity, we assume $N=M$ in the rest of the paper.